

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: LI, ET AL.
- (ii) TITLE OF INVENTION: Neurotransmitter Transporter
- (iii) NUMBER OF SEQUENCES: 2
- (iv) CORRESPONDENCE ADDRESS:
- (A) ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
CECCHI, STEWART & OLSTEIN
- (B) STREET: 6 BECKER FARM ROAD
- (C) CITY: ROSELAND
- (D) STATE: NEW JERSEY
- (E) COUNTRY: USA
- (F) ZIP: 07068
- (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: 3.5 INCH DISKETTE
- (B) COMPUTER: IBM PS/2
- (C) OPERATING SYSTEM: MS-DOS
- (D) SOFTWARE: WORD PERFECT 5.1
- (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
- (B) FILING DATE: SUBMITTED HEREWITH
- (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA
- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: FERRARO, GREGORY D.  
(B) REGISTRATION NUMBER: 36,134  
(C) REFERENCE/DOCKET NUMBER: 325800-118

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 201-994-1700  
(B) TELEFAX: 201-994-1744

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 2,486 BASE PAIRS  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: SINGLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGGAGGCAGG GAGTGAGGAG CGAGCGGAGT CGCGTGCGCC GGC GCGAGCT CCGGGTCGCC 60  
CCAGCCCCAG CCGGGGGCCT GTGGCGGGGG AGGAGCTGTG CGTCCGCGAC CCGTCGGGA 120  
TCG CAGCTGC TCGGCCGAG TGCACGGGCC GAGTCTGCGC GACTACCCAC GCGTGACAGG 180  
TCCCTGAATG AGAAGGAGCT GACAGCAGCT GAATTCCATC TTCTCTGTGT GCTGGGGAGC 240  
AGGGCTACAC GGCC CAGGTG GCATCAATGC CGAAGAACAG CAAAGTGACC CAGCGTGAGC 300  
ACAGCAGTGA GCATGTCACT GAGTCCGTGG CCGACCTGCT GGCCCTCGAG GAGCCTGTGG 360  
ACTATAAGCA GAGTG TACTG AATGTGGCTG GTGAGGCAGG CGGCAAGCAG AAGGCGGTGG 420  
AGGAGGAGCT GGATGCAGAG GACCGGCCGG CCTGGAACAG TAAGCTGCAG TACATCCTGG 480  
CCCAGATTGG CTTCTCTGTG GGCTCGGCA ACATCTGGAG GTTCCCCTAC CTGTGCCAGA 540  
AAAATGGAGG AGGTGCTTAC CTGGTGCCCT ACCTGGTGCT GCTGATCATC ATCGGGATCC 600  
CCCTCTTCTT CCTGGAGCTG GCTGTGGGTC AGAGGATCCG CCGCGGAAGC ATCGGTGTGT 660  
GGCACTATAT ATG TCCCCGC CTGGGGGGGA TCGGCTTCTC CAGCTGCATA GTCTGTCTCT 720  
TTGTGGGGCT GTATTATAAT GTGATCATCG GGTGGAGCAT CTTCTATTTT TC AAGTCCT 780  
TCCAGTACCC GCTGCCCTGG AGTGAATGTC CTGTCTGTCAG GAATGGGAGC GTCG CAGTGG 840  
TGGAGGCAGA GTGTGAAAAG AGCTCAGCCA CTACCTACTT CTGGTACCGA GAGGCTTTGG 900  
ACATCTCTGA CTCCATCTCG GAGAGTGGGG GCCTCAACTG GAAGATGACC CTGTGCCTCC 960  
TCGTGGTCTG GAGCATCGGG GGGATGGCTG TCGGTAAGGG CATCCAGTCC TCGGGGAAGG 1020

TGATGTATTT CAGCTCCCTC TTCCCCTACG TGGTGCTGGC CTGCTTCCTG GTCCGGGGGT 1080  
 TGTTGTTGCG AGGGGCAGTT GATGGCATCC TACACATGTT CACTCCCAAG CTGGTCAAGA 1140  
 TGCTGGACCC CCAGGTGTGG CGGGAGGTAG CTACCCAGGT CTTCCTTGGC TTGGGTCTGG 1200  
 GCTTTGGTGG TGTCATTGTC TTCTCCAGTT ACAATAAGCA GGACAACAAC TGCCACTTCG 1260  
 ATGGCGCCCT GGTGTCTTC ATCAACTTCT TCACGTCAGT GTTGGCCACC CTCGTGGTGT 1320  
 TTGTTGTTTT GGGCTTCAAG GCCAACATCA TGAATGAGAA GTGTGTGGTC GAGAATGCTG 1380  
 AGAAAATCCT AGGGTACCTT AACACCAACG TCCTGAGCCG GGACCTCATC CCACCCACG 1440  
 TCAACTTCTC CCACCTGACC ACAAAGGACT ACATGGAGAT GGACAATGTC ATCATGACCG 1500  
 TGAAGGAGGA CCAGTTCTCA GCCCTGGGCC TTGACCCCTG CCTTCTGGAG GACGAGCTGG 1560  
 ACAAGTCCGT GCAGGGCACA GGCCTGGCCT TCATCGCCTT CACTGAGGCC ATGACGCACT 1620  
 TCCCCACCTC CCCGTTCTGG TCCGTCATGT TCTTCTTGAT GCTTATCAAC CTGGGCCTGG 1680  
 GCAGCATGAT CGGGACCATG GCAGGCATCA CCACGCCCAT CATCGACACC TCCAAGGTGC 1740  
 CCAAGGAGAT GTTCACAGTG GGCTGCTGTG TCTTTACATT CCTCGTGGGA CTGTTGTTCTG 1800  
 TCCAGCGCTC CGGAAACTAC TTTGTCACCA TGTTGATGA CTACTCAGCC ACGCTGCCAC 1860  
 TCACTCTCAT CGTCATCCTT GAGAACATCG CTGTGGCCTG GATTTATGGA CCAAGAAGT 1920  
 TCATGCAGGA GCTGACGGAG ATGCTGGGCT TCCGCCCCTA CCGCTTCTAT TTCTACATGT 1980  
 GGAAGTTCGT GTCTCCACTA TGCATGGCTG TGCTCACCAC AGCCAGCATC ATCCAGCTGG 2040  
 GGGTCACGCC CCCGGCCTAC AGCGCCTGGA TCAAGGAGGA GGCTGCCGAG CGCTACCTGT 2100  
 ATTTCCCCAA CTGGCCCATG GCACTCCTGA TCACCCTCAT CGTCGTGGCG ACGCTGCCCA 2160  
 TCCCTGTGGT GTTCGTCCTG CGGCACTTCC ACCTGCTCTC TGATGGCTCC AACACCCTCT 2220  
 CCGTGTCTTA CAAGAAGGCC CGCATGATGA AGGACATCTC CAACCTGGAG GAGAACGATG 2280  
 AGACCCGCTT CATCTCAGC AAGGTGCCCA GTGAGGCACC TTCCCCCATG CCCACTCACC 2340  
 GTTCCTATCT GGGGCCCCGGC AGCACATCAC CCCTGGAGAC CAGCTGGAAC CCAATGGAC 2400  
 CCTATGGGCG CGGCTACCTG CTGGCCAGCA CCCCTGAGTC TGAGCTGTGA CCACTGCCCA 2460  
 AGCCCATGCC CGCTCTCCCC CCACCG 2486

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 727 AMINO ACIDS
- (B) TYPE: AMINO ACID
- (C) STRANDEDNESS:
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Pro Lys Asn Ser Lys Val Thr Gln Arg Glu His Ser Ser Glu

5

10

15

His Val Thr Glu Ser Val Ala Asp Leu Leu Ala Leu Glu Glu Pro	20	25	30
Val Asp Tyr Lys Gln Ser Val Leu Asn Val Ala Gly Glu Ala Gly	35	40	45
Gly Lys Gln Lys Ala Val Glu Glu Glu Leu Asp Ala Glu Asp Arg	50	55	60
Pro Ala Trp Asn Ser Lys Leu Gln Tyr Ile Leu Ala Gln Ile Gly	65	70	75
Phe Ser Val Gly Leu Gly Asn Ile Trp Arg Phe Pro Tyr Leu Cys	80	85	90
Gln Lys Asn Gly Gly Gly Ala Tyr Lys Val Pro Tyr Leu Val Leu	95	100	105
Leu Ile Ile Ile Gly Ile Pro Leu Phe Phe Leu Glu Leu Ala Val	110	115	120
Gly Gln Arg Ile Arg Arg Gly Ser Ile Gly Val Trp His Tyr Ile	125	130	135
Cys Pro Arg Leu Gly Gly Ile Gly Phe Ser Ser Cys Ile Val Cys	140	145	150
Leu Phe Val Gly Leu Tyr Tyr Asn Val Ile Ile Gly Trp Ser Ile	155	160	165
Phe Tyr Phe Phe Lys Ser Phe Gln Tyr Pro Leu Pro Trp Ser Glu	170	175	180
Cys Pro Val Val Arg Asn Glu Ser Val Ala Val Val Glu Ala Glu	185	190	195
Cys Glu Lys Ser Ser Ala Thr Thr Tyr Phe Trp Tyr Arg Glu Ala	200	205	210
Leu Asp Ile Ser Asp Ser Ile Ser Glu Ser Gly Gly Leu Asn Trp	215	220	225
Lys Met Thr Leu Cys Leu Leu Val Val Trp Ser Ile Gly Gly Met	230	235	240
Ala Val Gly Lys Gly Ile Gln Ser Ser Gly Lys Val Met Tyr Phe	245	250	255
Ser Ser Leu Phe Pro Tyr Val Val Leu Ala Cys Phe Leu Val Arg	260	265	270

Gly	Leu	Leu	Leu	Arg	Gly	Ala	Val	Asp	Gly	Ile	Leu	His	Met	Phe	275	280	285
Thr	Pro	Lys	Leu	Val	Lys	Met	Leu	Asp	Pro	Gln	Val	Trp	Arg	Glu	290	295	300
Val	Ala	Thr	Gln	Val	Phe	Phe	Gly	Leu	Gly	Leu	Gly	Phe	Gly	Gly	305	310	315
Val	Ile	Val	Phe	Ser	Ser	Tyr	Asn	Lys	Gln	Asp	Asn	Asn	Cys	His	320	325	330
Phe	Asp	Gly	Ala	Leu	Val	Ser	Phe	Ile	Asn	Phe	Phe	Thr	Ser	Val	335	340	345
Leu	Ala	Thr	Leu	Val	Val	Phe	Val	Val	Leu	Gly	Phe	Lys	Ala	Asn	350	355	360
Ile	Met	Asn	Glu	Lys	Cys	Val	Val	Glu	Asn	Ala	Glu	Lys	Ile	Leu	365	370	375
Gly	Tyr	Leu	Asn	Thr	Asn	Val	Leu	Ser	Arg	Asp	Leu	Ile	Pro	Pro	380	385	390
His	Val	Asn	Phe	Ser	His	Leu	Thr	Thr	Lys	Asp	Tyr	Met	Glu	Met	395	400	405
Asp	Asn	Val	Ile	Met	Thr	Val	Lys	Glu	Asp	Gln	Phe	Ser	Ala	Leu	410	415	420
Gly	Leu	Asp	Pro	Cys	Leu	Leu	Glu	Asp	Glu	Leu	Asp	Lys	Ser	Val	425	430	435
Gln	Gly	Thr	Gly	Leu	Ala	Phe	Ile	Ala	Phe	Thr	Glu	Ala	Met	Thr	440	445	450
His	Phe	Pro	Thr	Ser	Pro	Phe	Trp	Ser	Val	Met	Phe	Phe	Leu	Met	455	460	465
Leu	Ile	Asn	Leu	Gly	Leu	Gly	Ser	Met	Ile	Gly	Thr	Met	Ala	Gly	470	475	480
Ile	Thr	Thr	Pro	Ile	Ile	Asp	Thr	Ser	Lys	Val	Pro	Lys	Glu	Met	485	490	495
Phe	Thr	Val	Gly	Cys	Cys	Val	Phe	Thr	Phe	Leu	Val	Gly	Leu	Leu	500	505	510
Phe	Val	Gln	Arg	Ser	Gly	Asn	Tyr	Phe	Val	Thr	Met	Phe	Asp	Asp	515	520	525
Tyr	Ser	Ala	Thr	Leu	Pro	Leu	Thr	Leu	Ile	Val	Ile	Leu	Glu	Asn			

530	535	540
Ile Ala Val Ala Trp	Ile Tyr Gly Pro Lys Lys Phe Met Gln Glu	
545	550	555
Leu Thr Glu Met Leu	Gly Phe Arg Pro Tyr Arg Phe Tyr Phe Tyr	
560	565	570
Met Trp Lys Phe Val	Ser Pro Leu Cys Met Ala Val Leu Thr Thr	
575	580	585
Ala Ser Ile Ile Gln	Leu Gly Val Thr Pro Pro Ala Tyr Ser Ala	
590	595	600
Trp Ile Lys Glu Glu	Ala Ala Glu Arg Tyr Leu Tyr Phe Pro Asn	
605	610	615
Trp Pro Met Ala Leu	Leu Ile Thr Leu Ile Val Val Ala Thr Leu	
620	625	630
Pro Ile Pro Val Val	Phe Val Leu Arg His Phe His Leu Leu Ser	
635	640	645
Asp Gly Ser Asn Thr	Leu Ser Val Ser Tyr Lys Lys Ala Arg Met	
650	655	660
Met Lys Asp Ile Ser	Asn Leu Glu Glu Asn Asp Glu Thr Arg Phe	
665	670	675
Ile Leu Ser Lys Val	Pro Ser Glu Ala Pro Ser Pro Met Pro Thr	
680	685	690
His Arg Ser Tyr Leu	Gly Pro Gly Ser Thr Ser Pro Leu Glu Thr	
695	700	705
Ser Trp Asn Pro Asn	Gly Pro Tyr Gly Arg Gly Tyr Leu Leu Ala	
710	715	720
Ser Thr Pro Glu Ser	Glu Leu	
725		

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: LI, YI  
FLEISCHMANN, ROBERT
- (ii) TITLE OF INVENTION: NEUROTRANSMITTER TRANSPORTER
- (iii) NUMBER OF SEQUENCES: 6
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Human Genome Sciences, Inc.
  - (B) STREET: 9410 Key West Avenue
  - (C) CITY: Rockville
  - (D) STATE: MD
  - (E) COUNTRY: US
  - (F) ZIP: 20850
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/424,424
  - (B) FILING DATE: 21-APR-1995
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Michele M. Wales
  - (B) REGISTRATION NUMBER: 43,975
  - (C) REFERENCE/DOCKET NUMBER: PF116
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (301) 309-8504
  - (B) TELEFAX: (301) 309-8439

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2485 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 266..2446

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGGAGGCAGG GAGTGAGGAG CGAGCGGAGT CGCGTGCGCC GGCGCGAGCT CCGGGTCGCC 60  
 CCAGCCCCAG CCGGGGGCCT GTGGCGGGGG AGGAGCTGTG CGTCCGCGAC CCGTCGGGAT 120

CGCAGCTGCT CGGCCGGAGT GCACGGGCCG AGTCTGCGCG ACTACCCACG CGTGACAGGT	180
CCCTGAATGA GAAGGAGCTG ACAGCAGCTG AATTCCATCT TCTCTGTGTG CTGGGGAGCA	240
GGGCTACACG GCCCAGGTGG CATCA ATG CCG AAG AAC AGC AAA GTG ACC CAG Met Pro Lys Asn Ser Lys Val Thr Gln 1 5	292
CGT GAG CAC AGC AGT GAG CAT GTC ACT GAG TCC GTG GCC GAC CTG CTG Arg Glu His Ser Ser Glu His Val Thr Glu Ser Val Ala Asp Leu Leu 10 15 20 25	340
GCC CTC GAG GAG CCT GTG GAC TAT AAG CAG AGT GTA CTG AAT GTG GCT Ala Leu Glu Glu Pro Val Asp Tyr Lys Gln Ser Val Leu Asn Val Ala 30 35 40	388
GGT GAG GCA GGC GGC AAG CAG AAG GCG GTG GAG GAG GAG CTG GAT GCA Gly Glu Ala Gly Gly Lys Gln Lys Ala Val Glu Glu Glu Leu Asp Ala 45 50 55	436
GAG GAC CGG CCG GCC TGG AAC AGT AAG CTG CAG TAC ATC CTG GCC CAG Glu Asp Arg Pro Ala Trp Asn Ser Lys Leu Gln Tyr Ile Leu Ala Gln 60 65 70	484
ATT GGC TTC TCT GTG GGC CTC GGC AAC ATC TGG AGG TTC CCC TAC CTG Ile Gly Phe Ser Val Gly Leu Gly Asn Ile Trp Arg Phe Pro Tyr Leu 75 80 85	532
TGC CAG AAA AAT GGA GGA GGT GCT TAC CTG GTG CCC TAC CTG GTG CTG Cys Gln Lys Asn Gly Gly Gly Ala Tyr Leu Val Pro Tyr Leu Val Leu 90 95 100 105	580
CTG ATC ATC ATC GGG ATC CCC CTC TTC TTC CTG GAG CTG GCT GTG GGT Leu Ile Ile Ile Gly Ile Pro Leu Phe Phe Leu Glu Leu Ala Val Gly 110 115 120	628
CAG AGG ATC CGC CGC GGA AGC ATC GGT GTG TGG CAC TAT ATA TGT CCC Gln Arg Ile Arg Arg Gly Ser Ile Gly Val Trp His Tyr Ile Cys Pro 125 130 135	676
CGC CTG GGG GGG ATC GGC TTC TCC AGC TGC ATA GTC TGT CTC TTT GTG Arg Leu Gly Gly Ile Gly Phe Ser Ser Cys Ile Val Cys Leu Phe Val 140 145 150	724
GGG CTG TAT TAT AAT GTG ATC ATC GGG TGG AGC ATC TTC TAT TTC TTC Gly Leu Tyr Tyr Asn Val Ile Ile Gly Trp Ser Ile Phe Tyr Phe Phe 155 160 165	772
AAG TCC TTC CAG TAC CCG CTG CCC TGG AGT GAA TGT CCT GTC GTC AGG Lys Ser Phe Gln Tyr Pro Leu Pro Trp Ser Glu Cys Pro Val Val Arg 170 175 180 185	820
AAT GGG AGC GTC GCA GTG GTG GAG GCA GAG TGT GAA AAG AGC TCA GCC Asn Gly Ser Val Ala Val Val Glu Ala Glu Cys Glu Lys Ser Ser Ala 190 195 200	868
ACT ACC TAC TTC TGG TAC CGA GAG GCT TTG GAC ATC TCT GAC TCC ATC Thr Thr Tyr Phe Trp Tyr Arg Glu Ala Leu Asp Ile Ser Asp Ser Ile 205 210 215	916
TCG GAG AGT GGG GGC CTC AAC TGG AAG ATG ACC CTG TGC CTC CTC GTG Ser Glu Ser Gly Gly Leu Asn Trp Lys Met Thr Leu Cys Leu Leu Val 220 225 230	964



GTC TGG AGC ATC GGG GGG ATG GCT GTC GGT AAG GGC ATC CAG TCC TCG	1012
Val Trp Ser Ile Gly Gly Met Ala Val Gly Lys Gly Ile Gln Ser Ser	
235 240 245	
GGG AAG GTG ATG TAT TTC AGC TCC CTC TTC CCC TAC GTG GTG CTG GCC	1060
Gly Lys Val Met Tyr Phe Ser Ser Leu Phe Pro Tyr Val Val Leu Ala	
250 255 260 265	
TGC TTC CTG GTC CGG GGG TTG TTG TTG CGA GGG GCA GTT GAT GGC ATC	1108
Cys Phe Leu Val Arg Gly Leu Leu Leu Arg Gly Ala Val Asp Gly Ile	
270 275 280	
CTA CAC ATG TTC ACT CCC AAG CTG GTC AAG ATG CTG GAC CCC CAG GTG	1156
Leu His Met Phe Thr Pro Lys Leu Val Lys Met Leu Asp Pro Gln Val	
285 290 295	
TGG CGG GAG GTA GCT ACC CAG GTC TTC TTT GGC TTG GGT CTG GGC TTT	1204
Trp Arg Glu Val Ala Thr Gln Val Phe Phe Gly Leu Gly Leu Gly Phe	
300 305 310	
GGT GGT GTC ATT GTC TTC TCC AGT TAC AAT AAG CAG GAC AAC AAC TGC	1252
Gly Gly Val Ile Val Phe Ser Ser Tyr Asn Lys Gln Asp Asn Asn Cys	
315 320 325	
CAC TTC GAT GGC GCC CTG GTG TCC TTC ATC AAC TTC TTC ACG TCA GTG	1300
His Phe Asp Gly Ala Leu Val Ser Phe Ile Asn Phe Phe Thr Ser Val	
330 335 340 345	
TTG GCC ACC CTC GTG GTG TTT GTT GTT TTG GGC TTC AAG GCC AAC ATC	1348
Leu Ala Thr Leu Val Val Phe Val Val Leu Gly Phe Lys Ala Asn Ile	
350 355 360	
ATG AAT GAG AAG TGT GTG GTC GAG AAT GCT GAG AAA ATC CTA GGG TAC	1396
Met Asn Glu Lys Cys Val Val Glu Asn Ala Glu Lys Ile Leu Gly Tyr	
365 370 375	
CTT AAC ACC AAC GTC CTG AGC CGG GAC CTC ATC CCA CCC CAC GTC AAC	1444
Leu Asn Thr Asn Val Leu Ser Arg Asp Leu Ile Pro Pro His Val Asn	
380 385 390	
TTC TCC CAC CTG ACC ACA AAG GAC TAC ATG GAG ATG GAC AAT GTC ATC	1492
Phe Ser His Leu Thr Thr Lys Asp Tyr Met Glu Met Asp Asn Val Ile	
395 400 405	
ATG ACC GTG AAG GAG GAC CAG TTC TCA GCC CTG GGC CTT GAC CCC TGC	1540
Met Thr Val Lys Glu Asp Gln Phe Ser Ala Leu Gly Leu Asp Pro Cys	
410 415 420 425	
CTT CTG GAG GAC GAG CTG GAC AAG TCC GTG CAG GGC ACA GGC CTG GCC	1588
Leu Leu Glu Asp Glu Leu Asp Lys Ser Val Gln Gly Thr Gly Leu Ala	
430 435 440	
TTC ATC GCC TTC ACT GAG GCC ATG ACG CAC TTC CCC ACC TCC CCG TTC	1636
Phe Ile Ala Phe Thr Glu Ala Met Thr His Phe Pro Thr Ser Pro Phe	
445 450 455	
TGG TCC GTC ATG TTC TTC TTG ATG CTT ATC AAC CTG GGC CTG GGC AGC	1684
Trp Ser Val Met Phe Phe Leu Met Leu Ile Asn Leu Gly Leu Gly Ser	
460 465 470	
ATG ATC GGG ACC ATG GCA GGC ATC ACC ACG CCC ATC ATC GAC ACC TCC	1732
Met Ile Gly Thr Met Ala Gly Ile Thr Thr Pro Ile Ile Asp Thr Ser	

T030300 1111111111

475	480	485	
AAG GTG CCC AAG GAG ATG TTC ACA GTG GGC TGC TGT GTC TTT ACA TTC Lys Val Pro Lys Glu Met Phe Thr Val Gly Cys Cys Val Phe Thr Phe 490 495 500 505			1780
CTC GTG GGA CTG TTG TTC GTC CAG CGC TCC GGA AAC TAC TTT GTC ACC Leu Val Gly Leu Leu Phe Val Gln Arg Ser Gly Asn Tyr Phe Val Thr 510 515 520			1828
ATG TTC GAT GAC TAC TCA GCC ACG CTG CCA CTC ACT CTC ATC GTC ATC Met Phe Asp Asp Tyr Ser Ala Thr Leu Pro Leu Thr Leu Ile Val Ile 525 530 535			1876
CTT GAG AAC ATC GCT GTG GCC TGG ATT TAT GGA CCC AAG AAG TTC ATG Leu Glu Asn Ile Ala Val Ala Trp Ile Tyr Gly Pro Lys Lys Phe Met 540 545 550			1924
CAG GAG CTG ACG GAG ATG CTG GGC TTC CGC CCC TAC CGC TTC TAT TTC Gln Glu Leu Thr Glu Met Leu Gly Phe Arg Pro Tyr Arg Phe Tyr Phe 555 560 565			1972
TAC ATG TGG AAG TTC GTG TCT CCA CTA TGC ATG GCT GTG CTC ACC ACA Tyr Met Trp Lys Phe Val Ser Pro Leu Cys Met Ala Val Leu Thr Thr 570 575 580 585			2020
GCC AGC ATC ATC CAG CTG GGG GTC ACG CCC CCG GCC TAC AGC GCC TGG Ala Ser Ile Ile Gln Leu Gly Val Thr Pro Pro Ala Tyr Ser Ala Trp 590 595 600			2068
ATC AAG GAG GAG GCT GCC GAG CGC TAC CTG TAT TTC CCC AAC TGG CCC Ile Lys Glu Glu Ala Ala Glu Arg Tyr Leu Tyr Phe Pro Asn Trp Pro 605 610 615			2116
ATG GCA CTC CTG ATC ACC CTC ATC GTC GTG GCG ACG CTG CCC ATC CCT Met Ala Leu Leu Ile Thr Leu Ile Val Val Ala Thr Leu Pro Ile Pro 620 625 630			2164
GTG GTG TTC GTC CTG CGG CAC TTC CAC CTG CTC TCT GAT GGC TCC AAC Val Val Phe Val Leu Arg His Phe His Leu Leu Ser Asp Gly Ser Asn 635 640 645			2212
ACC CTC TCC GTG TCC TAC AAG AAG GCC CGC ATG ATG AAG GAC ATC TCC Thr Leu Ser Val Ser Tyr Lys Lys Ala Arg Met Met Lys Asp Ile Ser 650 655 660 665			2260
AAC CTG GAG GAG AAC GAT GAG ACC CGC TTC ATC CTC AGC AAG GTG CCC Asn Leu Glu Glu Asn Asp Glu Thr Arg Phe Ile Leu Ser Lys Val Pro 670 675 680			2308
AGT GAG GCA CCT TCC CCC ATG CCC ACT CAC CGT TCC TAT CTG GGG CCC Ser Glu Ala Pro Ser Pro Met Pro Thr His Arg Ser Tyr Leu Gly Pro 685 690 695			2356
GGC AGC ACA TCA CCC CTG GAG ACC AGC TGG AAC CCC AAT GGA CCC TAT Gly Ser Thr Ser Pro Leu Glu Thr Ser Trp Asn Pro Asn Gly Pro Tyr 700 705 710			2404
GGG CGC GGC TAC CTG CTG GCC AGC ACC CCT GAG TCT GAG CTG Gly Arg Gly Tyr Leu Leu Ala Ser Thr Pro Glu Ser Glu Leu 715 720 725			2446
TGACCACTGC CCAAGCCCAT GCCCGCTCTC CCCCCACCG			2485

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 727 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Pro	Lys	Asn	Ser	Lys	Val	Thr	Gln	Arg	Glu	His	Ser	Ser	Glu	His
1				5					10						15
Val	Thr	Glu	Ser	Val	Ala	Asp	Leu	Leu	Ala	Leu	Glu	Glu	Pro	Val	Asp
			20					25					30		
Tyr	Lys	Gln	Ser	Val	Leu	Asn	Val	Ala	Gly	Glu	Ala	Gly	Gly	Lys	Gln
		35					40					45			
Lys	Ala	Val	Glu	Glu	Glu	Leu	Asp	Ala	Glu	Asp	Arg	Pro	Ala	Trp	Asn
	50					55					60				
Ser	Lys	Leu	Gln	Tyr	Ile	Leu	Ala	Gln	Ile	Gly	Phe	Ser	Val	Gly	Leu
65					70					75					80
Gly	Asn	Ile	Trp	Arg	Phe	Pro	Tyr	Leu	Cys	Gln	Lys	Asn	Gly	Gly	Gly
				85					90					95	
Ala	Tyr	Leu	Val	Pro	Tyr	Leu	Val	Leu	Leu	Ile	Ile	Ile	Gly	Ile	Pro
			100					105						110	
Leu	Phe	Phe	Leu	Glu	Leu	Ala	Val	Gly	Gln	Arg	Ile	Arg	Arg	Gly	Ser
			115				120					125			
Ile	Gly	Val	Trp	His	Tyr	Ile	Cys	Pro	Arg	Leu	Gly	Gly	Ile	Gly	Phe
	130					135					140				
Ser	Ser	Cys	Ile	Val	Cys	Leu	Phe	Val	Gly	Leu	Tyr	Tyr	Asn	Val	Ile
145					150					155					160
Ile	Gly	Trp	Ser	Ile	Phe	Tyr	Phe	Phe	Lys	Ser	Phe	Gln	Tyr	Pro	Leu
				165					170					175	
Pro	Trp	Ser	Glu	Cys	Pro	Val	Val	Arg	Asn	Gly	Ser	Val	Ala	Val	Val
			180					185					190		
Glu	Ala	Glu	Cys	Glu	Lys	Ser	Ser	Ala	Thr	Thr	Tyr	Phe	Trp	Tyr	Arg
			195				200					205			
Glu	Ala	Leu	Asp	Ile	Ser	Asp	Ser	Ile	Ser	Glu	Ser	Gly	Gly	Leu	Asn
	210					215					220				
Trp	Lys	Met	Thr	Leu	Cys	Leu	Leu	Val	Val	Trp	Ser	Ile	Gly	Gly	Met
225					230					235					240
Ala	Val	Gly	Lys	Gly	Ile	Gln	Ser	Ser	Gly	Lys	Val	Met	Tyr	Phe	Ser
				245					250					255	
Ser	Leu	Phe	Pro	Tyr	Val	Val	Leu	Ala	Cys	Phe	Leu	Val	Arg	Gly	Leu
			260					265					270		

Leu	Leu	Arg	Gly	Ala	Val	Asp	Gly	Ile	Leu	His	Met	Phe	Thr	Pro	Lys	
		275					280									285
Leu	Val	Lys	Met	Leu	Asp	Pro	Gln	Val	Trp	Arg	Glu	Val	Ala	Thr	Gln	
		290				295					300					
Val	Phe	Phe	Gly	Leu	Gly	Leu	Gly	Phe	Gly	Gly	Val	Ile	Val	Phe	Ser	
305					310					315					320	
Ser	Tyr	Asn	Lys	Gln	Asp	Asn	Asn	Cys	His	Phe	Asp	Gly	Ala	Leu	Val	
				325					330					335		
Ser	Phe	Ile	Asn	Phe	Phe	Thr	Ser	Val	Leu	Ala	Thr	Leu	Val	Val	Phe	
			340					345					350			
Val	Val	Leu	Gly	Phe	Lys	Ala	Asn	Ile	Met	Asn	Glu	Lys	Cys	Val	Val	
		355					360					365				
Glu	Asn	Ala	Glu	Lys	Ile	Leu	Gly	Tyr	Leu	Asn	Thr	Asn	Val	Leu	Ser	
		370				375					380					
Arg	Asp	Leu	Ile	Pro	Pro	His	Val	Asn	Phe	Ser	His	Leu	Thr	Thr	Lys	
385					390					395					400	
Asp	Tyr	Met	Glu	Met	Asp	Asn	Val	Ile	Met	Thr	Val	Lys	Glu	Asp	Gln	
				405					410					415		
Phe	Ser	Ala	Leu	Gly	Leu	Asp	Pro	Cys	Leu	Leu	Glu	Asp	Glu	Leu	Asp	
			420					425					430			
Lys	Ser	Val	Gln	Gly	Thr	Gly	Leu	Ala	Phe	Ile	Ala	Phe	Thr	Glu	Ala	
		435					440					445				
Met	Thr	His	Phe	Pro	Thr	Ser	Pro	Phe	Trp	Ser	Val	Met	Phe	Phe	Leu	
		450				455					460					
Met	Leu	Ile	Asn	Leu	Gly	Leu	Gly	Ser	Met	Ile	Gly	Thr	Met	Ala	Gly	
465					470					475					480	
Ile	Thr	Thr	Pro	Ile	Ile	Asp	Thr	Ser	Lys	Val	Pro	Lys	Glu	Met	Phe	
				485					490					495		
Thr	Val	Gly	Cys	Cys	Val	Phe	Thr	Phe	Leu	Val	Gly	Leu	Leu	Phe	Val	
			500					505					510			
Gln	Arg	Ser	Gly	Asn	Tyr	Phe	Val	Thr	Met	Phe	Asp	Asp	Tyr	Ser	Ala	
			515				520					525				
Thr	Leu	Pro	Leu	Thr	Leu	Ile	Val	Ile	Leu	Glu	Asn	Ile	Ala	Val	Ala	
						535					540					
Trp	Ile	Tyr	Gly	Pro	Lys	Lys	Phe	Met	Gln	Glu	Leu	Thr	Glu	Met	Leu	
545					550					555					560	
Gly	Phe	Arg	Pro	Tyr	Arg	Phe	Tyr	Phe	Tyr	Met	Trp	Lys	Phe	Val	Ser	
				565					570					575		
Pro	Leu	Cys	Met	Ala	Val	Leu	Thr	Thr	Ala	Ser	Ile	Ile	Gln	Leu	Gly	
			580					585					590			
Val	Thr	Pro	Pro	Ala	Tyr	Ser	Ala	Trp	Ile	Lys						

Ser Thr Pro Glu Ser Glu Leu  
725

- (A) LENGTH: 38 base pairs

- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

38

(2). INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 29 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

29